An Ontology-Driven Application to Improve the Prescription of Educational Resources to Parents of Premature Infants

Howard Goldberg[1], Alfredo Morales[1], David MacMillan[2], and Matthew Quinlan[2]

[1] Clinician Support Technologies Inc, 1 Wells Avenue, Newton, MA 02459
[2] Network Inference Limited, 25 Chapel Street, London NW1 5DH

Abstract. CST's Baby CareLink provides a 'collaborative healthware' environment for parents of premature infants that incorporates just-in-time learning as one means of knowledge exploration and patient empowerment [1], [2], [3]. As the Baby CareLink content base has continued to grow, it has become increasingly difficult for content prescribers to identify all relevant resources for parents at a given point in time in their child's course of care. In addition, the growing content base has become increasingly difficult to maintain without a rich indexing system. In order to address these issues, we have developed an ontology-driven application that supports the indexing and retrieval of educational materials according to rich descriptions of premature infants.

We have developed an initial OWL-DL-based [4] ontology describing relevant concepts in the domain of neonatology, including clinical conditions, diagnostic testing, therapies, durable medical equipment, and the infants themselves. We have developed an initial terminologic model describing typical clinical problems and therapies that occur over the clinical course of these premature infants. Indexers tag educational resources through a web based client application that allows them to create rich descriptions of educational resources based on the reference ontology. Clinical end-users interact with a client application that identifies educational resources appropriate to clinical scenarios occurring over the course of a typical premature infant's development based on the description generated from records of existing infants in the Baby CareLink system, or according to a user-created description.

Network Inference's tools were used to develop the neonatology ontology and implement the run-time system. Construct™, a Visio-based ontology modeling tool was used to develop the reference ontology. This tool allowed the representation of the domain's concepts, subsumption relationships, properties, instances, and axioms diagrammatically. Cerebra Server™, an OWL-DL-based inferencing platform was used to provide logical consistency-checking at modeling time directly from Construct™. Cerebra Server™ was integrated with Baby CareLink and to the indexing and retrieving tools through a .Net connector that provides an API for 1) extending the ontology at indexing time with newly-classified educational resources, 2) dynamically creating instances of individual premature babies using data from Baby CareLink and 3) querying the ontology at run-time.

This knowledge centric approach to the identification and recommendation of educational resources is anticipated to better individualize information for parents of infants managed through Baby CareLink and increase the efficacy of the information prescription process. The approach is expected to increase the return on investment provided by Baby CareLink through reduction in the time required by nurses and care managers to interact with the system.

Introduction

CST's Baby CareLink provides a 'collaborative healthware' environment for parents of premature infants that incorporates just-in-time learning as one means of knowledge exploration and patient empowerment [1], [2], [3]. As the Baby CareLink content base has continued to grow, it has become increasingly difficult for content prescribers to identify all relevant resources for parents at a given point in time in their child's course of care. In addition, the growing content base has become increasingly difficult to maintain without a rich indexing system. In order to address these issues, we have developed an ontology-driven application that supports the indexing and retrieval of educational materials according to rich descriptions of premature infants.

Ontology Design

Initial work on the Proof of Concept (POC) focused on modelling the neonatology domain, representing educational resources, and designing appropriate queries. The resulting model represents a base ontology composed of two main types of constructs—hierarchies of core concepts and axioms describing prototypical premature infants. The resulting ontology comprises approximately 300 entities, including concepts, object properties, and complex concepts.

The foundation for the model represents the core domain concepts to be used within the ontology. The core concepts are organized in hierarchies describing babies, problems, treatments, tests and educational resources. Each category was elaborated to a level of granularity required in the context of the POC. For example, Treatment comprises Therapies, Medications and Procedures. Categories of Baby were defined according to standard properties, such as gestational age, using classes with customized complex datatypes to express value ranges within which patients could be classified. A number of Object Properties were defined (e.g. "may present", "may be treated", "may be tested") in order to facilitate the definition of relationships between concepts. These are also used later in the process of querying the live ontology

An additional class called 'Typical' was introduced into the model. The Typical class facilitates the creation of a taxonomic model describing prototypical premature infants, their presenting problems, and usual therapies. This strategy addresses reuse by permitting very general axioms in the model, e.g., "*A typical 24-36 week old* *premature infant presenting with respiratory distress syndrome may be treated with surfactant and either mechanical ventilation or CPAP*" without burdening all instantiations of 24-30 week-old babies with the generalization. In the application, we use this duality to enable caregivers to subsequently differentiate between educational resources relevant to those problems a baby might be **expected** to present and those based upon what the baby is **known** to present.

The model of prototypical premature infants contained in the ontology consists of complex concepts that define required and optional combinations of tests and treatments for problems presented by babies within specific age ranges. The OWL-DL constructs *intersectionOf* and *unionOf* were used to define relationships and restrictions over concepts in the ontology. Problems, Treatments and Tests were associated graphically with the intersection of the Typical class and sub-classes of Baby (see Figure 1).

Finally, the Educational Resource class was represented. Education Resources 'refer to' arbitrarily complex descriptions of premature infants. The representations used to index educational resources are created programmatically through end-user tools described in the following section.

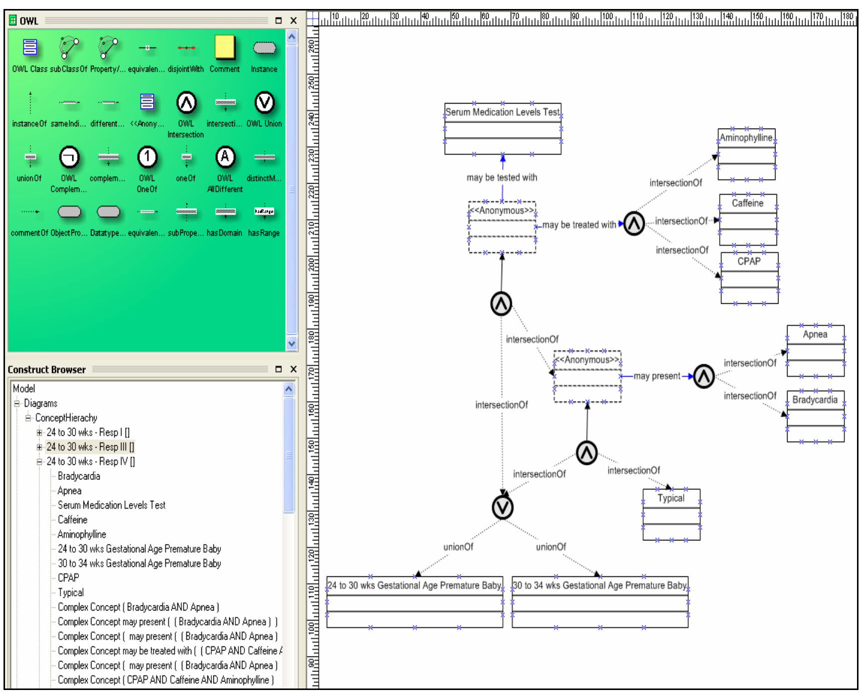

Fig. 1: Description of Babies with Gestational Age 24-34 weeks, in relation to specific problems, treatments, and tests

System Overview

The application is architected to integrate at author-time with Baby CareLink's content management system (CMS). Domain experts create and publish educational resources through a web-based editing workbench, part of the CMS. Two further tools were added to the workbench to support the description of resources—the Resource Descriptor Plug-in and the Publishing Wizard. Figure 2 shows the architectural components and their interactions at each stage.

The Resource Descriptor Plug-in allows domain experts to generate metadata about resources in the Baby CareLink content base, using templates that represent prototypical relationships between concepts and constructs that exist in the base ontology. The plug-in dynamically obtains concepts through XQueries posted to Cerebra via its client API. The outputs of the plug-in are resource descriptors in the form of OWL-DL fragments. Each resource descriptor is an instance of the Educational Resource concept, restricted by the combinations of Problems, Treatments and/or Tests relating to a rich description for a premature infant. The source for resource descriptors is stored in the content management system.

The Publishing Wizard facilitates the deployment of resource descriptors into Cerebra. It interacts with both the CMS and Cerebra to merge completed resource descriptors into the active ontology. The source for the resulting extended ontology must also be saved should the ontology need to be reloaded into Cerebra.

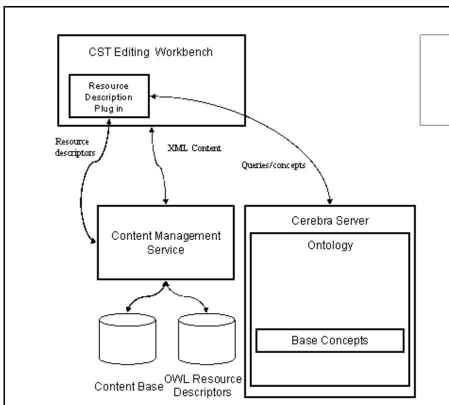

Figure 2.a – Domain expert creates resource descriptors using templates and concepts obtained from the ontology.

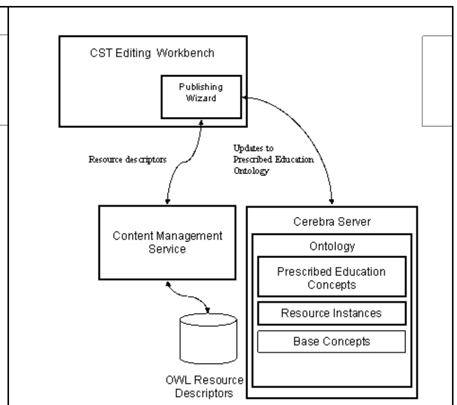

Figure 2.b – Once all education resources are described by resource descriptors, these are published to Cerebra.

In addition to the author-time tools, Baby CareLink was extended in order to represent patients within the inferencing system. A Patient Profiler module was developed that synthesizes CareLink data from registered infants into instances of the Baby concept. These instances serve as the focal point with which to identify educational resources for each patient. The Patient profiler acts as a daemon, executing at periodic intervals and forcing reclassification of the resulting ontology once new Baby instances have been merged.

At run-time, the Information Prescription Pad extends the existing Prescribed Education module by providing rapid, fine-grained searches against the universe of educational resources known to Cerebra. Clinicians interact with the Pad's user interface in order to create a description of an appropriate baby to be used for search. The interface presents a series of templates for completion using concepts from the ontology. Clinicians may build a description from an existing baby or a prototypical baby.

The Information Prescription Pad queries Cerebra in order to retrieve concepts that may be valid fillers for slots of the selected template. Posting an instanceProperty query to Cerebra using a Baby instance and an object property will return all valid fillers for the specified property. If the clinician is interested in what typically could be relevant to a particular patient, the Pad uses the specific Baby instance representing that patient as part of the query – i.e., the conjunction of 'Premature Baby 1' and the concept 'Typical'. If the clinician is interested in a prototypical baby, the Pad will use the concept that represents that type of baby when querying Cerebra Server – i.e., 'Premature Baby with 24 to 30 weeks Gestational Age' and the concept 'Typical'. Once all templates have been completed, the Information Prescription Pad formulates an XQuery and posts it to Cerebra through the client API.

The query returns instances of Educational Resources that match the criteria specified in the template. Each instance has a property containing the URI of the resource to be prescribed. The Information Prescription Pad renders the list of applicable resources and provides both a way to review the document and to assign it to the parents of a baby. After relevant resources have been chosen, the Information Prescription Pad updates the User Profile of the chosen parent, assigning the resources for her review the next time that she accesses Baby CareLink.

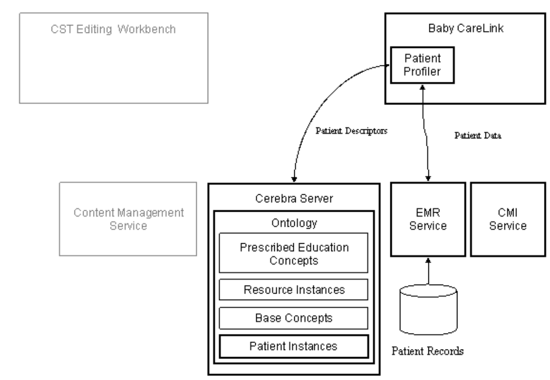

Fig. 3.a. Baby CareLink creates instances describing patients and merges them into Cerebra.

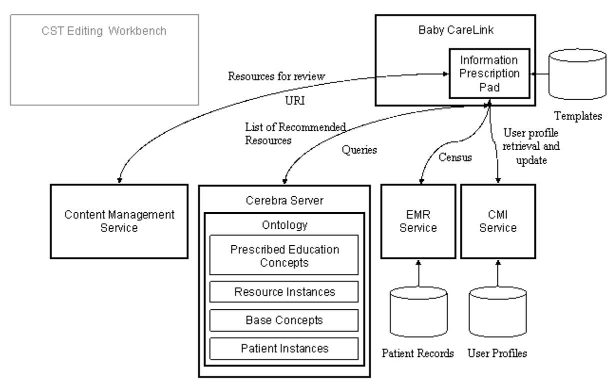

Fig. 3.b. Clinicians use the Information Prescription Pad to query Cerebra for educational resources

Evaluation

At the time of submission, the POC system is being alpha tested at CST. Domain experts are in the process of defining resource descriptors for the more than 800 documents that comprise the Baby CareLink content base, using the Resource Descriptor Plug-in added to the Editing Workbench. The first stage of the process centers on creating metadata for the subset of documents related to respiratory problems, their treatments and the tests that could be performed on a baby presenting such problems. The domain experts' feedback is being used to refine the templates originally developed.

Conclusions and Impacts

We have developed an ontology-driven application that facilitates the prescription of educational materials to parents of premature infants. Our ontology supports reasonably high-fidelity representations of neonates, their clinical problems, and ongoing treatments. These representations allow for a rapid, fine-grained search against a document set of approximately eight hundred documents. The ontology has also supported the development of a taxonomic model of potential problems and treatments occurring in typical neonates over time. The taxonomic model supports the creation of a module to answer "what if" questions regarding typical clinical scenarios, e.g., "What are the typical treatments for a 27-week-only baby with respiratory distress syndrome?"

This early work is significant in several dimensions. We were able to develop a small, but robust ontology supporting a sophisticated retrieval application using the current draft of OWL-DL. The feature set of the Cerebra Server provided model-time validation of the developing ontology, real-time updates to the ontology, and adequate querying and inferencing capabilities for both terminologic reasoning, and reasoning about instances. We are able to incorporate Cerebra as a modular inferencing service within a larger software architecture.

While the CST modelers had previous familiarity with description logic systems such as LOOM [5], they found that OWL-DL still had a significant learning curve. For example, understanding OWL representations for necessary and sufficient conditions requires multiple re-readings of the OWL reference manual to ensure correct subsumption relationships. The Construct modeling tool simplifies this by hiding the verbosity of the OWL language, but the modelers also found it was necessary to think in terms of axiomatic representations in addition to object models. Much of the collaborative effort between CST and Network Inference was spent in identifying critical OWL modeling idioms that would scale as the ontology grew. Additionally, while terminologic query capabilities were adequate for this application, there remains ground for additional improvements in the expressiveness of a query language for OWL. Standard notions of time, which are important for our application, remain to be adapted for an OWL environment. Finally, while we were able to adapt datatypes for use in characterizing age ranges for our infant models, it would be useful to incorporate numeric comparison operators into OWL as well.

Future work will examine the use or integration of existing healthcare ontologies into our system. Generalization of our work to additional medical domains is best accomplished through extension of existing ontologies as opposed to *de novo* development. Fortunately, there is an existing base of formal ontologies for the health care domain, such as GALEN [6] and SNOMED [7], with which to develop applications. The features and employed idioms of these ontologies must be evaluated to establish their use in knowledge mediators such as the one we have developed. Additionally, these are very large ontologies, whose applicability to Description Logic reasoning and performance characteristics must be determined for the inference engine and architecture we are deploying.

Our early work adds additional evidence to the utility of ontology-driven knowledge mediators, as previously demonstrated by systems such as TAMBIS [8] and Ariadne/SIMS [9]. This work adds the additional capability for the system to reason over the domain of interest through the additional terminologic model of typical premature infants over time. By addressing real-world implementations of ontology-driven knowledge mediators, we believe this class of mediators may be one of the early 'killer applications' for semantic web technologies.

References

1. Gray JE, Safran C, Davis RB, Pompilio-Weitzner G, Stewart JE, Zaccagnini L, Pursley D.: Baby CareLink: Using the Internet and Telemedicine to Improve Care for High-Risk Infants. In: Pediatrics. (2000) Dec; 106(6):1318-24.
2. Goldberg H, Morales A, Gottlieb L, Meador L, Safran C. Reinventing Patient-Centered Computing for the Twenty-first Century. Medinfo 2001, (2001) 1455- 1458
3. Safran, C. The Collaborative Edge: Patient Empowerment for Vulnerable Populations. International Journal of Medical Informatics 69 (2003) 185-190.
4. McGuinness, D. L., van Harmelen, F.: OWL Web Ontology Language Overview W3C Working Draft. W3C (2003).
5. MacGregor RM. Using a Description Classifier to Enhance Deductive Inference. *Proceedings Seventh IEEE Conference on AI Applications*, pp. 141-147, 1991.
6. Rector AL, Bechhofer SK, Goble CA, Horrocks I, Nowlan WA, Solomon WD. The GRAIL Concept Modelling Language for Medical Terminology**.** *Artificial Intelligence in Medicine,* Volume 9, 1997.
7. Spackman KA, Campbell KE, Cote RA. SNOMED RT: A reference terminology for health care. *Proceedings/AMIA Annual Fall Symposium*:640-4, 1997.
8. Goble CA, Stevens R, Ng G, Bechhofer S, Paton NW, Baker PG, Peim M, and Brass A. Transparent Access to Multiple Bioinformatics Information Sources. IBM Systems Journal, 40(2):532 - 552, 2001.
9. Knoblock CA. Planning, executing, sensing, and replanning for information gathering. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Montreal, Canada, 1995.